ROBUST DELAY-&-PREDICT EQUALIZATION FOR BLIND SIMO CHANNEL DEREVERBERATION

Mahdi Triki[†]*, and Dirk T.M. Slock*[*]

[†]Digital Signal Processing Group, Philips Research Laboratories

High Tech Campus 36, Eindhoven, The Netherlands

[*] Eurecom Institute, 2229 route des Crêtes, B.P. 193, 06904 Sophia Antipolis Cedex, France

Email: mahdi.triki@philips.com, dirk.slock@eurecom.fr

ABSTRACT

We[1] consider the blind multichannel dereverberation problem for a single source. We have shown before [5] that the single-input multi-output (SIMO) reverberation filter can be equalized blindly by applying multivariate Linear Prediction (LP) to its output (after SISO input pre-whitening). In this paper, we investigate the LP-based dereverberation in a noisy environment, and/or under acoustic channel length underestimation. Considering ambient noise and late reverberation as additive noises, we propose to introduce a postfilter that transforms the multivariate prediction filter into a somewhat longer equalizer. The postfilter allows to equalize to non-zero delay. Both MMSE-ZF and MMSE design criteria are considered here for the postfilter. Simulations show that the proposed scheme is robust in noisy environments and channel length underestimation, and performs better compared to the classic Delay-&-Predict equalizer and the Delay-&-Sum beamformer.

1. INTRODUCTION

Blind dereverberation is the process of removing the effect of reverberation from an observed reverberant signal. Reducing the distortion caused by reverberation is a difficult blind deconvolution problem, due to the colored and non-stationary nature of speech and the length of the equivalent impulse response from the speaker's mouth to the microphone(s). Consider a clean speech signal, s_k, produced in a reverberant room. The reverberant speech signal observed on M distinct microphones can be written as:

$$\mathbf{y}_k = \mathbf{h}(q)\, s_k \tag{1}$$

where $\mathbf{y}_k = [y_{1,k} \cdots y_{M,k}]^T$ is the reverberant speech signal, $\mathbf{h}(z) = [h_1(z) \cdots h_M(z)]^T = \sum_{i=0}^{L_h-1} \mathbf{h}_i z^{-i}$ is the SIMO FIR channel transfer function, L_h is the channel length. The introduction of q, where q^{-1} is the one sample time delay operator: $q^{-1}s_k = s_{k-1}$, allows to introduce the compact notation of transfer functions in the time domain (whereas z in the z-transform is a complex number).

Blind dereverberation faces the channel/speech source identifiability problem. In fact, for any invertible scalar filter $\alpha(q)$, $(\alpha(q)\mathbf{h}(q), (1/\alpha(q))\, s_k)$ is also an acceptable solution for (1). In [1], the authors compute a multichannel FIR equalizer using a subspace based method. The identifiability problem is solved using accurate information of the "source" (or "noise") subspace dimension. The validity of the technique hinges critically on the true channel impulse response being of strictly finite duration, and its successful identification requires knowledge of the channel length [2]. For the acoustic case, the true channel impulse response length is generally unknown and/or ill-defined. This is a major limitation to the practical applicability of the subspace based methods to speech dereverberation.

In contrast, the alternative Linear Prediction (LP) based technique (proposed and refined by Slock et al. [3, 4]) proved to be consistent in the presence of channel order error. This makes the LP equalizer one of the more attractive solutions for blind speech dereverberation, as proposed in [5]. One tricky issue though is that in order for the LP to perform zero delay channel equalization, the source should be white, otherwise LP will perform both channel equalization and source whitening. Hence, in the case of speech dereverberation, some additional processing is required. In [5, 6], the speech correlation gets compensated via a SISO pre-whitening at the LP equalizer input (microphone signals). Next, the multivariate LP can be computed, and applied to the reverberant microphone signals $\mathbf{y}_k$:

$$\underbrace{\mathbf{u}_k}_{M\times 1} = \underbrace{A(q)}_{M\times M}\,\underbrace{\mathbf{y}_k}_{M\times 1} = \underbrace{\mathbf{h}_0}_{M\times 1}\,\underbrace{s_k}_{1\times 1} \quad \text{since} \quad A(z)\,\mathbf{h}(z) = \mathbf{h}_0 \tag{2}$$

where $A(q)$ is the MIMO linear prediction error filter, and $\mathbf{h}_0 = \mathbf{h}(z{=}+\infty)$ is the multichannel precursor coefficient. The LP equalizer is obtained by performing Maximum Ratio Combining (MRC) ($\mathbf{h}_0^T$) on the prediction error signal $\mathbf{u}_k$ components.

In [8] a somewhat related approach has been proposed, in which only the first microphone signal (assumed to have the shortest delay) is predicted in terms of the past samples on all microphones (MISO prediction). Compared to MIMO prediction, MISO prediction loses the MRC advantage. Since the MISO prediction is applied directly to $\mathbf{y}_k$, a dereverberated but also whitened source signal gets produced. Now, multivariate channel prediction assumes that the individual microphone channel transfer functions $h_i(z)$ $(i = 1, \ldots, M)$ have no SISO transfer function factor in common. If such a common factor exists, or equivalently if the source is colored, the multivariate LP will model this factor with an all-pole filter and the LP filter will contain a scalar transfer function factor that is the inverse of the all-pole model. This scalar factor can be determined as the common roots of the M MISO LP component transfer function polynomials or, as in [8], as the eigenvalues of a large matrix of which the MISO LP coefficients constitute one column. Postfiltering of the MISO LP residual with the inverse of the extracted factor then allows to recover in principle the unwhitened source. This common root extraction approach is prone to ill-conditioning, as the results in [8] tend to confirm. Indeed, due to the tapered off behavior of the late reverberation on all microphones, the $h_i(z)$ tend to have zeros that cluster near the origin and hence that are close or (almost) in common. This is not a big problem for the MIMO LP approach in [5, 6] where the effect is that

[1]Eurecom Institute's research is partially supported by its industrial members: BMW, Bouygues Télécom, Cisco Systems, France Télécom, Hitachi Europe, SFR, Sharp, STMicroelectronics, Swisscom, Thales. The work reported herein was also supported by The European FP6 NoE project K-Space.

the reverberation tail will not get equalized, but it is small anyway. For the purpose of the determination of the source color as in [8] on the other hand, the effect of such ill-conditioning is more severe.

In [9],[10] the so-called TRINICON method was introduced for blind separation of acoustic sources. One of the main characteristics of the objective function optimized by the TRINICON method, which is also based completely on second-order statistics (SOS), is that the extracted sources at the output of a MIMO FIR filter are as jointly decorrelated as possible, apart from intra-source correlations. In other words, the MIMO FIR demixing filter tries to produce source estimates with as little inter-source correlation as possible. As a result the cascade of the MIMO demixing and mixing filters will tend to a diagonal MIMO filter (apart from source permutations) and hence the sources may appear in a filtered fashion. Hence the problem solved is not so much that of dereverberation but of source separation. Also, the method is only applicable starting with at least two sources. And in spite of being SOS based, the objective function is not quadratic and requires an iterative (natural gradient based) solution.

Dereverberation techniques are generally introduced in a noiseless environment (the problem is already quite difficult even under these ideal conditions). In this paper, we propose a robust scheme for dereverberation in the presence of noise. This noise may be either additive acoustic noise or residual late reverberation due to underestimation of the reverberation delay spread (for computational complexity reasons or for estimation considerations in non-stationary environments). We investigate the resulting dereverberation performance in both a noisy environment and under the impulse response length underestimation.

We next summarize the basic D-&-P equalization technique from [5, 6, 7]. At first $\mathbf{y}_k$ gets replaced by $D(q)\,\mathbf{y}_k$, a microphone-wise delayed version of the microphone signals so that the source signal arrives with the same delay at all microphones. We shall denote the aligned version of $\mathbf{y}_k$ still by $\mathbf{y}_k$. Next, a SISO source LP filter $A_s(z)$ gets determined by performing LP on the $y_{i,k}$ SOS averaged over the M microphones. We then obtain $\mathbf{x}_k = A_s(q)\,\mathbf{y}_k = \mathbf{h}(q)\,\widetilde{s}_k$ where $\widetilde{s}_k = A_s(q)\,s_k$ is the whitened source signal. MIMO LP on $\mathbf{x}_k$ yields a prediction error

$$\widetilde{\mathbf{x}}_k = A_{\mathbf{x}}(q)\,\mathbf{x}_k = \mathbf{h}_0\,\widetilde{s}_k \;\text{ with }\; A_{\mathbf{x}}(z)\,\mathbf{h}(z) = \mathbf{h}_0\,. \tag{3}$$

Finally, the dereverberated source gets estimated as $\widehat{s}_k = \mathbf{h}_0^T\,A_{\mathbf{x}}(q)\,\mathbf{y}_k$.

2. ROBUST DELAY-&-PREDICT EQUALIZATION IN NOISY ENVIRONMENTS

In a noisy environment, the microphone signals can be written as

$$\mathbf{y}_k = \mathbf{h}(q)\,s_k + \mathbf{v}_k \tag{4}$$

where the noise $\mathbf{v}_k$ represents acoustic noise and/or the effect of modeling error in $\mathbf{h}(z)$. We shall model $\mathbf{v}_k$ as spatiotemporally white noise, independent of s_k. Such noise, for given noise power, is the worst case noise. In any case, at medium to high SNR, the correlation of the noise is a secondary effect compared to accounting for the noise power. The SISO and MIMO LP problems in the dereverberation approach considered here should still be formulated for the noise-free signals, even in the noisy case. However, since the LP problems only involve SOS, the noiseless SOS can easily be obtained from the noisy SOS in the white noise hypothesis, especially in the multichannel configuration considered here in which signal and noise subspaces arise. The simplest SOS denoising would be to subtract the noise covariance matrix ($\sigma_v^2 I$) from the covariance matrix $R_{\mathbf{y}}$ of $\mathbf{y}_k$ by estimating σ_v^2 from the noise subspace eigenvalue(s) of $R_{\mathbf{y}}$. Various degrees of sophistication are possible, that we shall not elaborate on here. Applying the (noiseless) MIMO LP to the noisy microphone signals, we get

$$\mathbf{u}_k = A_{\mathbf{x}}(q)\,\mathbf{y}_k = \mathbf{h}_0\,s_k + A_{\mathbf{x}}(q)\,\mathbf{v}_k\,. \tag{5}$$

The robustified D&P equalizer then gets constructed as

$$F_{D-\&-P}(q) = \mathbf{w}(q)\,A_{\mathbf{x}}(q)\,,\;\; \widehat{s}_k = F_{D-\&-P}(q)\,\mathbf{y}_k = \mathbf{w}(q)\,\mathbf{u}_k \tag{6}$$

whereas the basic D&P equalizer uses $\mathbf{w}(q) = \mathbf{h}_0^T$, which maximizes the power of the desired signal part but not necessarily the output SNR. In [7], we have proposed the postfilter $\mathbf{w}(q)$ with a MMSE-ZF design using explicitely the white noise hypothesis (in a multichannel configuration, there is an infinity of zero-forcing designs, one of which will be MMSE). The filter length of $\mathbf{w}(q)$ allows the design of non-zero-delay equalizers. Here we shall consider the design of the postfilter using the MMSE-ZF and MMSE criteria, without a white noise hypothesis.

MMSE-ZF Design

For a given filter length L_w and an equalization delay $0 \le d \le (L_w-1)$, the weighting filters are optimized by maximizing the output SNR (under the d-delay zero-forcing constraint), i.e.

$$\begin{cases} \mathbf{w} = \arg\max\limits_{\mathbf{w}} \dfrac{\sigma_s^2}{\oint \mathbf{w}(z) S_{\mathbf{u}}(z) \mathbf{w}^{\dagger}(z) \dfrac{dz}{2\pi j z} \;- \sigma_s^2} \\ \mathbf{w}(z)\,\mathbf{h}_0 = z^{-d} \end{cases} \tag{7}$$

where $\mathbf{w}^{\dagger}(z)$ denotes the paraconjugate (matched filter) of $\mathbf{w}(z)$, and $S_{\mathbf{u}}(z) = A_{\mathbf{x}}(z)\,S_{\mathbf{y}}(z)\,A_{\mathbf{x}}^{\dagger}(z)$ is the matrix spectrum of $\mathbf{u}_k$. For a time domain formulation, let $\underline{\mathbf{w}} = [\mathbf{w}_0 \cdots \mathbf{w}_{L_w-1}]$, $\mathbf{U}_k = [\mathbf{u}_k^T \cdots \mathbf{u}_{k-L_w+1}^T]^T$, $\mathbf{H}_0 = I_{L_w} \otimes \mathbf{h}_0$ and $\mathbf{e}_d = [0 \ldots 0\;1\;0 \ldots 0]$ with a 1 in position $d{+}1$. Hence $\widehat{s}_k = \underline{\mathbf{w}}\,\mathbf{U}_k$. The optimization in (7) becomes

$$\begin{cases} \underline{\mathbf{w}}_{L_w,d}^{zf} = \arg\min\limits_{\underline{\mathbf{w}}} \underline{\mathbf{w}}\,\mathbf{R}_{\mathbf{U}}\underline{\mathbf{w}}^T \\ \underline{\mathbf{w}}\,\mathbf{H}_0 = \mathbf{e}_d \end{cases} \tag{8}$$

where $\mathbf{R}_{\mathbf{U}}$ is short for $\mathbf{R}_{\mathbf{UU}} = E\,\mathbf{U}_k\mathbf{U}_k^T$, the covariance matrix of $\mathbf{u}_k$ of (block) size L_w. The optimal postfilter is

$$\underline{\mathbf{w}}_{L_w,d}^{zf} = \mathbf{e}_d \left(\mathbf{H}_0^T\mathbf{R}_{\mathbf{U}}^{-1}\mathbf{H}_0\right)^{-1}\mathbf{H}_0^T\mathbf{R}_{\mathbf{U}}^{-1} \tag{9}$$

with corresponding optimal

$$\mathrm{SNR}_{L_w,d}^{zf} = \frac{\sigma_s^2}{\mathbf{e}_d \left(\mathbf{H}_0^T\mathbf{R}_{\mathbf{U}}^{-1}\mathbf{H}_0\right)^{-1}\mathbf{e}_d^T \;- \sigma_s^2}\;. \tag{10}$$

The optimal delay (maximum SNR) corresponds to the position of the smallest diagonal element of $\left(\mathbf{H}_0^T\mathbf{R}_{\mathbf{U}}^{-1}\mathbf{H}_0\right)^{-1}$.

MMSE Design

The MMSE design corresponds to $\underline{\mathbf{w}}_{L_w,d}^{mmse} = \mathbf{R}_{q^{-d_s}\mathbf{U}}\mathbf{R}_{\mathbf{UU}}^{-1}$. Now $\mathbf{R}_{q^{-d_s}\mathbf{U}} = \mathbf{e}_d\,R_{SS}\,\mathbf{H}_0^T$ where R_{SS} is the source covariance matrix of size L_w, to be constructed using an AR model using the SISO LP filter. Note that $\mathbf{e}_d\,R_{SS}$ means that only row $d{+}1$ of R_{SS} needs to be computed. Hence $\underline{\mathbf{w}}_{L_w,d}^{mmse} = \mathbf{e}_d\,R_{SS}\,\mathbf{H}_0^T\mathbf{R}_{\mathbf{U}}^{-1}$ and

$$\mathrm{SNR}_{L_w,d}^{mmse} = \frac{\sigma_s^2}{\mathbf{e}_d\,R_{SS}\,\mathbf{H}_0^T\mathbf{R}_{\mathbf{U}}^{-1}\mathbf{H}_0\,R_{SS}\,\mathbf{e}_d^T} - 1\;. \tag{11}$$

3. EXPERIMENTAL RESULTS

3.1. MMSE-ZF postfiltering for robust dereverberation in noisy environment

We first illustrate the behavior of zero-forcing post-processing, and we provide a comparison with the classic Delay-&-Predict equalizer. We consider a rectangular room with dimensions $L_x = 8\ m$, $L_y = 10\ m$ and $L_z = 4\ m$, and with wall reflection coefficients $\rho_x = \rho_y = \rho_z = 0.9$ ($T_{60} = 250\ ms$). A speech signal with duration of 8.8s, and sampled at 8 kHz is used as the original source signal. The reverberant speech signal is observed on 2 distinct microphones. A computer implementation (graciously provided by Geert Rombouts while at K.U. Leuven) of the image method as described in [11] is used to generate synthetic room impulse responses for the microphones. We constrain the postfilter length (and hence the equalization delay d) to $L_w \leq 100$ ($d \leq 12.5\ ms$). The optimal delay (maximizing (10)) is selected. Figure 1 plots the Signal-to-

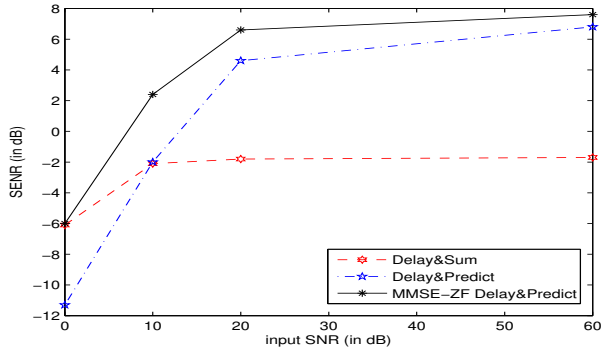

Fig. 1. The SENR function of the input SNR.

Echo+Noise Ratio (SENR $= \dfrac{\sum_k s_k^2}{\sum_k (s_k - \widehat{s}_k)^2}$) as a function of the input Signal-to-Noise Ratio (SNR $= \dfrac{\sum_k \|\mathbf{y}_k - \mathbf{v}_k\|^2}{\sum_k \|\mathbf{v}_k\|^2}$). The curves show that, in all regions, the MMSE-ZF D-&-P performs better than both the classic D-&-P and D-&-S. Particularly in a noisy environment, the postfiltering becomes essential in order to have acceptable enhancement accuracy. On the other hand, one can also remark that the post-processing still has a positive effect even in absence of ambient noise (SNR=60 dB). The reason is that the postfiltering also compensates for the errors in the estimation of the source spectrum (the estimation is done by averaging only two observation spectra ($M = 2$)).

3.2. MMSE-ZF Delay-&-Predict equalization under channel length underestimation

Ambient noise is not the unique source of additive noise. In fact, acoustic reverberation is theoretically infinitely long. As we assume that the channel has a finite length L_h, the late reverberation will be considered as additive noise, i.e.,

$$\mathbf{y}_k = \sum_{i=0}^{L_h-1} \mathbf{h}_i s_{k-i} + \underbrace{\sum_{i=L_h}^{\infty} \mathbf{h}_i s_{k-i}}_{\mathbf{v}_k}. \qquad (12)$$

Classically the channel length is chosen long enough such that the energy of the remaining reverberation is negligible (typically $L_h \geq T_{60} f_s$). With such a choice, the acoustic channels may have considerable length in real propagation environments. Hence, the algorithm may become computationally very expensive. In this section, we investigate the effect of undermodeling of the reverberation response on the dereverberation performance.

We model the late reverberation as a spherically diffuse noise [12] (although strictly speaking this additive noise (late reverberation) is neither white nor independent from the reverberant signal). Then, we apply MMSE-ZF postfiltering to reduce the late reverberation effect. We consider the Direct to Reverberant energy Ratio (DRR) as an evaluation criterion for the dereverberation accuracy:

$$\text{DRR} = 10 \log_{10} \left\{ \frac{\sum_{t=0}^{\tau-1} \widetilde{h}_t^2}{\sum_{t=\tau}^{\infty} \widetilde{h}_t^2} \right\} \ \text{dB} \qquad (13)$$

where $\widetilde{h}_t = \mathbf{f}_t * \mathbf{h}_t = \sum_i \mathbf{f}_i \mathbf{h}_{t-i}$ denotes the equalized channel (with a given equalizer $\mathbf{f}(q)$), and τ is the number of samples to be included as the direct component. The choice of the parameter τ depends on the application (how annoying early and late reverberation are in the given application). By increasing the value of τ, we give more weight to the degradation due to the late reverberation. If τ is small ($\tau \leq 1\ ms$), the DRR criterion will be correlated with the dereverberation SENR (equal if the input is white). Figures 2 and 3 plot the curves of the output DRR of the classic, MMSE-ZF Delay-&-Predict equalizers, and the Delay-&-Sum beamformer (function of the assumed channel length), respectively using 2 and 4 microphone array setup (for $\tau = 10\ ms$ and $\tau = 1\ ms$).

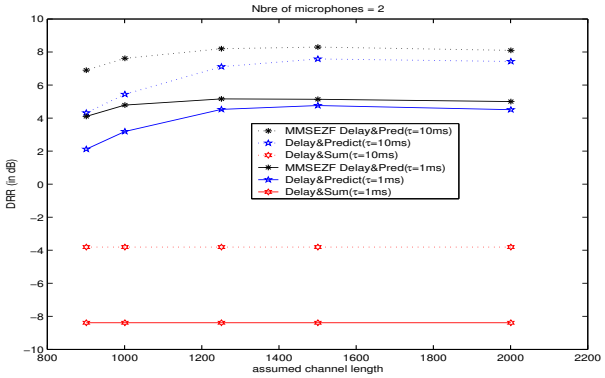

Fig. 2. The output DRR as function of the assumed channel length, using a 2 microphone array setup ($\tau = 10\ ms$ and $\tau = 1\ ms$)).

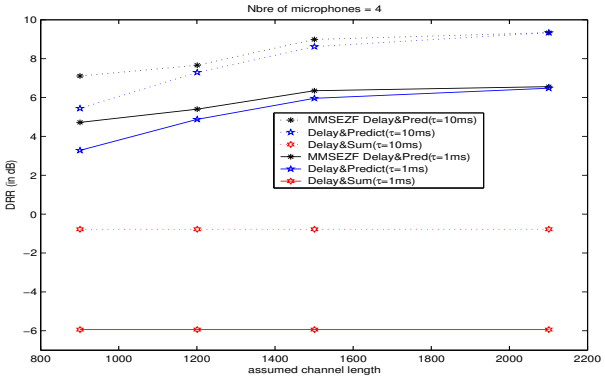

Fig. 3. The output DRR as function of the assumed channel length, using a 4 microphone array setup ($\tau = 10\ ms$ and $\tau = 1\ ms$).

One can remark that the MMSE-ZF D-&-P outperforms the classic

D-&-P in terms of equalization accuracy, and increases the robustness to channel length undermodeling. In all cases, the two schemes (classic and MMSE-ZF D-&-P) outperform the D-&-S beamformer. Also note that even when the channel length is over-estimated, the MMSE-ZF D-&-P still performs better than the classic D-&-P, especially when only few microphones are available. As stated in the previous section, this is due to the fact that the MMSE-ZF D-&-P can compensate for the errors in the estimation of the source correlations. These errors become more severe as the number of microphones decreases.

3.3. MMSE postfiltering for robust dereverberation

Next, we investigate the behavior of MMSE postfiltering, and we provide a comparison with the MMSE-ZF design. The source correlations R_{ss} are reconstructed from the estimated source AR model (as described in [5]). We consider the effect of both additive white noise and channel length underestimation. The postfilter lengths are constrained to $L_w \leq 100$. The same equalization delay is used for both MMSE and MMSE-ZF post-processing and is set to $d = 50$. As an evaluation criterion, we consider the SENR Gain (we consider the classic Delay-&-Sum performance as reference). We plot the SENR Gain as a function of the assumed channel length. The input SNR is set to 5 dB (figure 4) and 15 dB (figure 5), respectively.

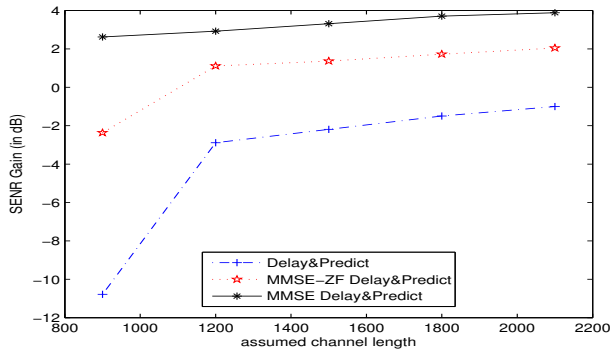

Fig. 4. The SENR Gain as function of the assumed channel length (SNR $= 5dB$, $M = 4$).

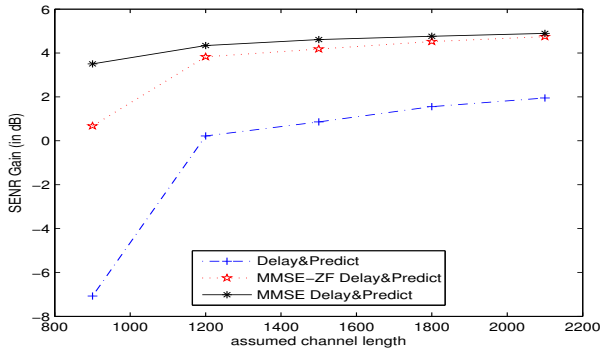

Fig. 5. The SENR Gain as function of the assumed channel length (SNR $= 15dB$, $M = 4$).

These simulations confirm that allowing for an equalization delay improves the overall dereverberation performance, and it is essential at low SNR. We observe also that MMSE postfiltering outperforms the MMSE-ZF based scheme, especially in the low SNR region. As expected, the benefit of an MMSE design vanishes at high SNR. One may remark that classic D-&-P may be inferior to D-&-S in these unideal conditions, but the new designs are precisely designed to handle those conditions.

4. CONCLUSIONS

In this paper, we have introduced robust Delay-&-Predict equalization for blind SIMO dereverberation. We have optimized the transformation of the multivariate prediction filter to a longer equalizer using the MSE criterion. The optimization is performed with or without zero-forcing constraints, leading respectively to MMSE-ZF and MMSE designs. The filter length increase allows for the introduction of some equalization delay, that can also be optimized. Experimental results illustrate that considerable gains can be achieved by allowing for a small equalization delay. It has also been shown that the post-processing is crucial in the low SNR region, increases robustness to the channel length underestimation and alleviates errors in the source color estimation. In these regions of interest, simulations prove that the MMSE design is more appropriate.

5. REFERENCES

[1] A. Assa-El-Bey, K. Abed-Meraim, and Y. Grenier. "Blind Separation of Audio Sources Convolutive Mixtures Using Parametric Decomposition," *In Proc of IWAENC*, Sept. 2005.

[2] A.J. van der Veen, S. Talwar, A. Paulraj. "A Subspace Approach to Blind Space-Time Signal Processing for Wireless Communication Systems," *IEEE Trans. on Acoustics, Speech and Signal Processing*, Vol.45, pp. 173-190, Jan. 1997.

[3] D.T.M. Slock. "Blind Fractionally-Spaced Equalization, Perfect-Reconstruction Filter Banks and Multichannel Linear Prediction," *In Proc. of IEEE ICASSP*, Apr. 1994.

[4] C.B. Papadias, and D.T.M. Slock. "Fractionally Spaced Equalization of Linear Polyphase Channels and Related Blind Techniques Based on Multichannel Linear Prediction," *IEEE Trans. on Signal Processing*, pp.641-654, Mar. 1999.

[5] M. Triki and D.T.M. Slock. "Blind Dereverberation of Quasi-periodic Sources Based on Multichannel Linear Prediction," *In Proc. of IWAENC*, Sept. 2005.

[6] M. Triki and D.T.M. Slock. "Delay and Predict Equalization For Blind Speech Dereverberation," *In Proc. of IEEE ICASSP*, Vol.5, pp.97-100, May 2006.

[7] M. Triki and D.T.M. Slock. "Multivariate LP Based MMSE-ZF Equalizer Design Considerations and Application to Multi-Microphone Dereverberation," *In Proc. of IEEE ICASSP*, Apr. 2007.

[8] M. Delcroix, T. Hikichi, M. Miyoshi. "Precise Dereverberation Using Multichannel Linear Prediction," *IEEE Trans. on Audio, Speech, and Language Processing*, Feb. 2007.

[9] H. Buchner, R. Aichner, W. Kellermann. "Blind Source Separation for Convolutive Mixtures: a Unified Treatment," in J. Benesty, Y. Huang (Eds.), *Audio Signal Processing for Next-generation Multimedia Communication Systems*, pp. 255-293, Kluwer Academic Publishers, Boston, MA, 2004.

[10] H. Buchner, R. Aichner, W. Kellermann. "A Generalization of Blind Source Separation Algorithms for Convolutive Mixtures based on Second-Order Statistics," *IEEE Trans. on Audio, Speech, and Language Processing*, pp. 120-134, Jan. 2005.

[11] P.M. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," *J. Acoust. Soc. Amer.*, pp.1527-1529, Nov. 1986.

[12] A. Koul, J.E. Greenberg, "Using Intermicrophone Correlation to Detect Speech in Spatially Separated Noise," *EURASIP Journal on Applied Signal Processing*, Issue 12, 2006.